Reviewer #3 (Anonymous)

Thank you for the chance to review this reflection on a recent ecohydrological summer school learning experience.

There is much in this paper to enjoy and draw on usefully. I do think that any models of innovative ways to deliver course content are valuable for the community to share, and I am fully supportive of the authors sharing these experiences in the HESS Special Issue.

I have a several concerns with the manuscript in its current form, however. Please treat these criticisms as a basis for improvement, rather than an attack on the motivation or the course and paper. I am focusing on these issues because they provide opportunities to improve the manuscript and focus the message. However, having just scanned my fellow reviewer's comments, it seems that they are shared concerns.

We are glad that the reviewer enjoyed this study and supports its inclusion in this HESS special issue on education. Also, we appreciate the criticisms offered up (which echo comment made by the other reviewers) as they have helped us focus our presentation and message.

In brief:
1. My most significant concern is that I am not convinced that the authors can robustly support their conclusions about the student orientation to active learning tasks based on the data they have collected.
2. I question just how ecohydrological this course really was – I see almost zero ecology/plant physiology in the course material.
3. The authors may be significantly under-estimating some of the challenges associated in scaling their approach to more typical classroom environments.
4. In terms of editorial issues – there are numerous spelling errors, typos, grammatical errors etc that need to be addressed. I've identified a few of them, but the manuscript is in need of a hard edit.

These concerns are elaborated on below.

We address these central concerns in response to the elaborations that follow.

1. Do the data support the conclusions?

There is a need to provide a more critical analysis of the student feedback and responses to inform future course planning. For future consideration, I would suggest that when working with such a small group of students, a different method of evaluation based on semi-structured qualitative interview techniques will be much more appropriate and useful than written evaluations. This approach is widely used in the educational literature to "go deeper" when sample sizes are small, so that even when there is limited statistical power, there is still scope to identify interesting educational outcomes based on student reflection. In this methodology, a neutral party (ideally an education researchers) offers a set of predefined interview questions, but has the opportunity to follow up on interesting points students make in depth. The interviews are

recorded, and are analyzed through group-based scoring methods to reduce selection bias etc by the main researchers.

The comment is an excellent suggestion on how to improve our (or any study's) ability to evaluate teaching methods as we target educational research. We have highlighted in the revision (in the Methodology) that such evaluation design would be useful and pointed out that we have not used it in the current study as a potential shortcoming. While this is a potential weakness of the current study, there is, as highlighted by this and other reviewers, much content that reminas valuable from the evaluation considered. By including this criticism of the methodolgy, we feel that we provide useful insight to readers of this HESS special issue.

In the absence of the more detailed information about the students' experiences that such a technique might have yielded, I just cannot agree with the authors that the increase in the evaluations along the trajectory of the 3 tasks can be attributed to the % of active learning. We can posit an almost infinitely large number of alternative hypotheses for the observed increase, and have no basis to dismiss them.

Alternative interpretation 1: Time.
Summer schools take time to "ramp up". Relationships need to be built. Instructors need to work out how to work with their student cohort. My experience in this (from the NSF Hydrological Synthesis Summer Institutes) is that the student teams become much more confident, comfortable and efficient as the summer schools progress. This leads to a more satisfying experience, and presumably more successful educational outcomes, later in the programs.

Alternative interpretation 2: Goals
Could it just be that the goals of the exercises were progressively more interesting to the students? The goal of the first task is pretty abstract. I do not think the authors have the data to discriminate between the effects of the teaching method and the teaching goals when looking at student responses. Even if we could be convinced that the teaching method was not optimal, how can we be sure that this is because of the degree of how "active" it was, versus e.g. students' familiarity with the scientific literature? Could a course of lectures ("passive") have actually achieved the results in a better way?

Alternative interpretation 3: Quality of task design
Given that several students seemed to find Task 1 poorly defined… could it just be that more thought and planning had gone into Tasks 2 and 3?

I could go on. My point is that the evaluations can't be taken as evidence that student learning scaled with how active the tasks were. There is no control, and the student feedback itself points to alternative explanations. I think this discussion must be revised in much more cautious terms.

We take this point to heart and have, thus, highlighted the potential for confounding influences on this study due to the limitations of the evaluation design and small sample size. Further, we have taken up more catious terms in the discussion through edits and inclusion of more critical assessment (as also suggested by other reviewers) of both our own methodology and the utility of active learning envirnonments.

2. How ecohydrological was this course really?

I refer primarily to Table 3 in the manuscript – the questions that are posed have nary a mention of a plant, a root depth, a stomata, or a crop type in there! While I agree that the models put together for predicting soil moisture balance & ET (a la Laio et al 2001, Feddes models, Porporato and co.) really relate 2 physical entities (soil moisture & evaporation flux), the thing that makes them ecohydrological is that they (i) account for stomatal closure, (ii) account for root depth. I just don't see much beyond micrometeorology & maybe specifically agricultural micrometeorology in Table 3. Could the tasks not be rephrased to at least try to look for evidence of plant water stress…or something that puts the plants in the picture? Ok, this is a criticism of the course rather than the paper, but given all the nice discussion about trans and cross-disciplinarity, it was a bit disappointing to see such a very traditional micromet approach to this task. Maybe the authors could reflect on ways to get the vegetation into the research questions (e.g. via porometry measurements, or comparing 2 diff. plants under the same irrigation regime, or something!) for future courses?

This comment reflects other review concerns and, in part, some of the teachers' concerns with the first offering of this course. We have addressed this in the revised manuscript through the explicit introduction of a section on getting this course to be more 'ecohydrological' in nature. In addition, with respect the the student-generated questions in Table 3, this specifically highlights some of the difficulties that are faced when incorporating purely active learning techniques into ecohydrology teaching. We have expanded the discussion around these aspect.

Further, we greatly appreciate this reviewer's detailed suggestions and will consider including them in future course offerings! We have, as such, also included these (and other considerations) in our revision.

3. Scaling up to the classroom

One thing I was confused about was the student:teacher ratio during this course. At worst it must have been 1:6, and I suspect that it was at times higher than that. If a teacher is a good teacher, then the personalized attention these students must have received surely partly drives the very positive response the students had to the course? Could you achieve these sorts of outcomes with 30 students and 1 instructor?

How dependent was the success of this course on having the students focus singly on the course topic for 4 weeks? Would it have survived intact in a "normal" curriculum situation where students time and attention would have been otherwise divided? What was the total time commitment students put in? Again looking at previous experiences at summer institutes, many students put in over 60 hours a week – which would come to 16 a week in a 15 week semester – a high course load, at least in the context I'm familiar with (undergrad and grad education in the USA).

On page 9344 the authors describe the importance of giving students raw data. My experience with undergraduates (a different context, I acknowledge), has been that raw data were a big problem. If students are not adept at working with large datasets then performing QA/QC on a dataset is a major task for them, and consumes energy that would be better spent on the

hydrological problem. I'm personally in 2 minds about it – I think it is important for students to have an appreciation of the effort and techniques involved in data preparation – but it was also a bummer to find students who had spent days of effort on it with limited success. Since this is a potential "trap" for folks looking to emulate your approach, perhaps some caution is needed?

We address these comments in our revision by adding the clear corollary that the teachers' confidence in scaling up the course is only valid for a summer course or a full-time course where students dedicate to the course full time. Different consideration would be needed with regard to recreating this course, for example, within the context of a standard schedule of courses. Regardless, while the high student:teacher ratio may have helped in giving a successful course, we feel that the impact of including active learning into our teaching would carry over to larger courses (with a potential trade-off of increased difficulty due to logistics).

In addition, with respect to having students work with raw data, we stand by our initial statement with regards to its importance but acknowledge the reviewer's concern here and have, thus, softened our initial statement. For this specific course, the potential dangers of having the students work with the raw data were out weighted by the teachers' a priori knowledge of student skill sets. This knowledge came about from previous interaction with the students and the prerequisites for the course. We have explicitly highlighted this in the revised text.

4. Editorial issues

Page 9341 – typo in sentence 2 (line 2-3).
Corrected

Page 9342 – typo in line 23 (reads synthesis, should read synthesize) and 24 (reads "Greek" should, I think, read "Greece")
Corrected

Page 9343 – lines 6-9 – took me several reads to understand the intention of the sentence. Maybe consider rewording?
Corrected

Page 9344 lines 1-3 – did you compile a list of these "teachable moments"? It might have been instructive…?
We have added several examples of the teachable moments that came up in this TLA. These teaching moments included, for example, comparison of potential versus actual evapotranspiration conceptualizations, discussion of plant transpiration/water uptake responses under drought conditions, and basic review of the differences between empirical and physics-based modeling approaches.

Page 9344 line 4 – what data were in this dataset!
We have listed that this dataset included temperature, precipitation and streamflow data

Page 9344 – did the students also have an opportunity to compare the estimates to observed data? It seems to me that in the absence of some empirical measure of ET (e.g. from a flux tower, from water balance closure, from sap flux, from soil moisture balances…) that this might have been a

slightly unsatisfying experience – 400mm worth of variation in estimates, but no sense of what the actual errors were??
We did not have access to ET flux tower data to help confirm estimates. It was possible to roughly confirm estimates using the experimental data collected on site in TLA #3.

Page 9345 – if you're going to discuss the "location's unique features" … perhaps you could share those unique features with your readers?
We have listed the uniqueness of the site which include its location and the proximity of agriculture, native, and recreational vegetation (landscaping and turf grass) under various management strategies.

Section 3 – perhaps title this as "Assessment of educational effectiveness: Methods" or something that makes it clear that the methods you're discussing here relate to the way that you assessed the effectiveness of the techniques, and not the methods used to implement the techniques (which you just described!).
We have renamed this section.

Page 9348 – line 3 – I think you mean cognizant rather than cognitive? Excuse the US spelling, my spell checker doesn't think cogniscent (which I think is the UK version, but wouldn't bet on) is a word!
Sentence changed in response to a previous comment.

Figures

Fig 1 – caption doesn't read well – poor grammar. Can you rephrase?
Corrected

Fig 2 – should read "effective" not "affective" (affective would mean "relating to the emotions", and I hope is not what was being asked!)
Corrected